

SPARSE RECOVERY VIA ORTHOGONAL LEAST-SQUARES UNDER PRESENCE OF NOISE

Abolfazl Hashemi and Haris Vikalo

Department of Electrical and Computer Engineering
University of Texas at Austin, Austin, TX, USA

ABSTRACT

We consider the Orthogonal Least-Squares (OLS) algorithm for the recovery of a m -dimensional k -sparse signal from a low number of noisy linear measurements. The Exact Recovery Condition (ERC) in bounded noisy scenario is established for OLS under certain condition on nonzero elements of the signal. The new result also improves the existing guarantees for Orthogonal Matching Pursuit (OMP) algorithm. In addition, This framework is employed to provide probabilistic guarantees for the case that the coefficient matrix is drawn at random according to Gaussian or Bernoulli distribution where we exploit some concentration properties. It is shown that under certain conditions, OLS recovers the true support in k iterations with high probability. This in turn demonstrates that $\mathcal{O}(k \log m)$ measurements is sufficient for exact recovery of sparse signals via OLS.

Index Terms— compressed sensing, orthogonal least-squares, sparse reconstruction, exact recovery condition, orthogonal matching pursuit

1. INTRODUCTION

In many practical scenarios signal of interest can be modeled as a sparse solution to an underdetermined linear systems of equations. Examples include sparse linear regression [1], compressed sensing [2], sparse channel estimation in communication systems [3,4], compressive DNA microarrays [5] and a number of other applications in signal processing and machine learning [6–8]. Consider the linear measurement model

$$\mathbf{y} = \mathbf{H}\boldsymbol{\beta} + \boldsymbol{\eta}, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^n$ denotes the vector of observations, $\mathbf{H} \in \mathbb{R}^{n \times m}$ is the coefficient matrix (i.e., a collection of features) assumed to be full rank, $\boldsymbol{\eta} \in \mathbb{R}^n$ is the additive observation noise vector, and $\boldsymbol{\beta} \in \mathbb{R}^m$ is a vector known to have at most k non-zero components (i.e., k is the sparsity level of $\boldsymbol{\beta}$). Finding a sparse approximation to $\boldsymbol{\beta}$ leads to a cardinality-constraint optimization problem. In particular, we would like to solve the so-called l_0 -constrained least-squares

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{H}\boldsymbol{\beta}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_0 \leq k. \quad (2)$$

The number of possible locations of non-zero entries in $\boldsymbol{\beta}$ scales combinatorially with n which renders 2 computationally challenging; in fact, the problem is NP-hard.

Accelerated approaches to find an approximate solution include a number of iterative heuristics that attempt to solve 2 greedily by identifying columns of \mathbf{H} which correspond to non-zero components of $\boldsymbol{\beta}$ via locally optimal decisions. Among the greedy methods, orthogonal matching pursuit (OMP) algorithm [9] has attracted particular attention in recent years. There have been numerous modifications of OMP proposed in the literature that enhance the performance of OMP [10–12]. The principal idea in these methods is to select multiple “good” indices in each iteration in order to reduce the cost of identification step and recover the true support in fewer iterations. Performance of OMP is evaluated in numerous scenarios and necessary and sufficient conditions for exact reconstruction are provided [13–18].

Orthogonal Least-Squares (OLS) method [19] has drawn attention in recent years [20–22] and its provable performance is analyzed in some scenarios. In [20] OLS is analyzed under the Exact Recovery Condition (ERC) for noiseless setup. Herzet et al. [23] provided coherence-based conditions for sparse recovery of signals via OLS when nonzero components obey some decay. In [24], sufficient conditions for exact recovery from noise-free data is stated when a subset of optimal indices is available. However, the existing analysis and performance guarantees for OLS are predominantly limited to the case of non-random measurements in noise-free setting.

In this paper, we establish ERC-based guarantees for OLS and show that given certain SNR conditions, OLS recovers location of nonzero elements of $\boldsymbol{\beta}$ in first k iteration. Furthermore, a probabilistic result is provided for the case of random measurements with l_2 -bounded additive noise where coefficient matrix \mathbf{H} is drawn at random from Gaussian or Bernoulli distribution. Specifically, we find a lower bound on the probability of sparse recovery in k iterations when the nonzero element of unknown vector with smallest magnitude satisfies certain condition. Consequently, we demonstrate that with $\mathcal{O}(k \log m)$ measurements OLS will succeed with probability arbitrarily close to one.

The rest of the paper is organized as follows. Section 2 reviews the preliminaries. In section 3, we establish sufficient conditions on exact support recovery from perturbed

measurements. Section 4 provides our results on performance guarantee of OLS from linear random measurements. Some concluding remarks are provided in Section 5.

2. PRELIMINARIES AND NOTATIONS

In this paper, Bold capital letters refer to matrices and bold lowercase letters represent vectors. For matrix $\mathbf{H} \in \mathbb{R}^{n \times m}$ with full column rank, i.e., $m > n$, \mathbf{H}_{ij} denotes the entry in the i^{th} row and column j^{th} , \mathbf{a}_j refers to the j^{th} column of \mathbf{H} , and $\mathbf{H}_k \in \mathbb{R}^{n \times k}$ is one of the $\binom{m}{k}$ submatrices of \mathbf{H} . $\mathcal{L}_{\mathbf{H}}$ is the subspace spanned by columns of \mathbf{H} . $\mathbf{P}_{\mathbf{H}}^{\perp} = \mathbf{I} - \mathbf{H}\mathbf{H}^{\dagger}$ is the orthogonal projection operator onto orthogonal complement of $\mathcal{L}_{\mathbf{H}}$ where $\mathbf{H}^{\dagger} = (\mathbf{H}^{\top}\mathbf{H})^{-1}\mathbf{H}^{\top}$ is the Moore-Penrose pseudo-inverse of \mathbf{H} and \mathbf{I} is the identity matrix whose dimension is equal to the number of rows in \mathbf{H} . Similar notations are defined for \mathcal{L}_k , \mathbf{H}_k^{\dagger} , and \mathbf{P}_k where we drop the subscript \mathbf{H} for simplicity. Specifically, \mathbf{P}_k is the projection operator onto \mathcal{L}_k .

Let $\mathcal{I} = \{1, \dots, m\}$ be the set of all indices, $\mathcal{S}_{\text{opt}} = \{1, \dots, k\}$ be the set of indices corresponding to nonzero elements of β , and \mathcal{S}_i be the set of selected indices at the end of i^{th} iteration of OLS. For the sets $\mathcal{T}_1 \subset \mathcal{I}$ and $\mathcal{T}_2 \subset \mathcal{I}$ define $\mathbf{b}_j^{\mathcal{T}_1} = \frac{\mathbf{P}_{\mathcal{T}_1}^{\perp} \mathbf{a}_j}{\|\mathbf{P}_{\mathcal{T}_1}^{\perp} \mathbf{a}_j\|_2}$, $j \in \mathcal{T}_2$ where $\mathbf{P}_{\mathcal{T}_1}^{\perp}$ is the orthogonal projection matrix onto orthogonal complement of subspace spanned by columns of \mathbf{H} with indices in \mathcal{T}_1 . $\mathbb{1}\{\cdot\}$ is the indicator function and is equal to 1 if its argument holds and zero otherwise.

For a scalar random variable X , the notation $X \sim \mathcal{B}(\frac{1}{2}, \pm 1)$ denotes that X is a Bernoulli random variable and takes values 1 and -1 with equal probability. For non-scalar object such as matrix \mathbf{H} , $\mathbf{H} \sim \mathcal{N}(0, \frac{1}{n})$ means entries of \mathbf{H} are drawn independently according to a zero-mean Gaussian distribution with variance $\frac{1}{n}$. Similar definition holds for $\mathbf{H} \sim \mathcal{B}(\frac{1}{2}, \pm \frac{1}{\sqrt{n}})$.

We recall the principles of the OLS algorithm. OLS sequentially projects columns of \mathbf{H} onto a residual vector and selects the column that leads to the smallest residual norm. Specifically, OLS chooses a new index j_s in i^{th} iteration by employing the following criterion:

$$j_s = \arg \min_{j \in \mathcal{I} \setminus \mathcal{S}_{i-1}} \left\| \mathbf{y} - \mathbf{H}_{\mathcal{S}_{i-1} \cup \{j\}} \mathbf{H}_{\mathcal{S}_{i-1} \cup \{j\}}^{\dagger} \mathbf{y} \right\|_2 \quad (3)$$

This procedure is computationally more expensive than OMP since in addition to solving a least-square problem to update the residual vector, orthogonal projection of each column needs to be found at each step of OLS. Note that the performances of OLS and OMP are identical when the columns of \mathbf{H} are orthogonal.¹ However, for coherent and Redundant dictionaries OLS outperforms OMP. See [24] for a detailed

¹In fact, orthogonality of the columns of \mathbf{A} leads to a modular objective function in 2, implying optimality of both methods when $\eta = 0$.

discussion It is shown in [25] that index selection criterion in 3 can alternatively be written as

$$j_s = \arg \max_{j \in \mathcal{I} \setminus \mathcal{S}_{i-1}} \left| \mathbf{r}_{i-1}^{\top} \frac{\mathbf{P}_{i-1}^{\perp} \mathbf{a}_j}{\|\mathbf{P}_{i-1}^{\perp} \mathbf{a}_j\|_2} \right| \quad (4)$$

where \mathbf{r}_{i-1} is the residual vector in i^{th} iteration. In addition, projection matrix needed for subsequent iteration is related to the current projection matrix by the following recursion,

$$\mathbf{P}_{i+1}^{\perp} = \mathbf{P}_i^{\perp} - \frac{\mathbf{P}_i^{\perp} \mathbf{a}_{j_s} \mathbf{a}_{j_s}^{\top} \mathbf{P}_i^{\perp}}{\|\mathbf{P}_i^{\perp} \mathbf{a}_{j_s}\|_2^2}. \quad (5)$$

Before formalizing the main results, we start by some useful lemmas that are employed in the proofs of main theorems.

Lemma 2.1. *Let \mathbf{H}_1 , \mathbf{H}_2 , and \mathbf{C} be full rank tall matrices such that $\mathbf{C} = [\mathbf{H}_1, \mathbf{H}_2]$. Then for $i = 1, 2$*

$$\sigma_{\min}(\mathbf{H}_i) \geq \sigma_{\min}(\mathbf{C}), \quad \sigma_{\max}(\mathbf{H}_i) \leq \sigma_{\max}(\mathbf{C}) \quad (6)$$

Lemma 2.2. *The noise term in 1 can equivalently be written as*

$$\boldsymbol{\eta} = \bar{\mathbf{H}} \mathbf{w} + \boldsymbol{\eta}^{\perp} \quad (7)$$

where $\boldsymbol{\eta}^{\perp} = \mathbf{P}_k^{\perp} \boldsymbol{\eta}$, $\mathbf{w} = \bar{\mathbf{H}}^{\dagger} \boldsymbol{\eta}$, and $\bar{\mathbf{H}}$ is a submatrix with all true columns of \mathbf{H} . In addition, for $(i+1)^{\text{st}}$ iteration

$$\mathbf{r}_i = \boldsymbol{\eta}^{\perp} + \mathbf{P}_i^{\perp} \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c} \quad (8a)$$

$$\|\mathbf{r}_i\|_2^2 = \|\boldsymbol{\eta}^{\perp}\|_2^2 + \|\mathbf{P}_i^{\perp} \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_2^2 \quad (8b)$$

where $\mathbf{z} = \bar{\beta} + \mathbf{w}$, $\bar{\beta}$ corresponds to nonzero elements of β , and subscript i^c denotes the set of optimal columns that have not been chosen in first i iterations.

Lemma 2.3. *Assume $\mathbf{H} \sim \mathcal{N}(0, 1/n)$ or $\mathbf{H} \sim \mathcal{B}(\frac{1}{2}, \pm \frac{1}{\sqrt{n}})$. Let $\mathbf{H}_k \in \mathbb{R}^{n \times k}$ be a submatrix of \mathbf{H} . Then, $\forall \mathbf{u} \in \mathbb{R}^n$ statistically independent of \mathbf{H}_k drawn according to $\mathbf{u} \sim \mathcal{N}(0, 1/n)$ or $\mathbf{u} \sim \mathcal{B}(\frac{1}{2}, \pm \frac{1}{\sqrt{n}})$, it holds that $\mathbb{E} \|\mathbf{P}_k \mathbf{u}\|_2^2 = \frac{k}{n} \mathbb{E} \|\mathbf{u}\|_2^2$. In addition, let $c_0(\epsilon) = \frac{\epsilon^2}{4} - \frac{\epsilon^3}{6}$. Then,*

$$\Pr\left\{ \left| \|\mathbf{P}_k \mathbf{u}\|_2^2 - \frac{k}{n} \mathbb{E} \|\mathbf{u}\|_2^2 \right| \leq \epsilon \frac{k}{n} \mathbb{E} \|\mathbf{u}\|_2^2 \right\} \geq 1 - 2e^{-kc_0(\epsilon)}. \quad (9)$$

3. EXACT RECOVERY CONDITION FOR OLS

The first analysis of OMP is due to Tropp [14] where he provided sufficient conditions for exact recovery of OMP in noise-free setting. Specifically, let $\tilde{\mathbf{H}} \in \mathbb{R}^{n \times k}$ be a matrix with columns indexed by \mathcal{S}_{opt} and $\bar{\mathbf{H}} \in \mathbb{R}^{n \times (m-k)}$ correspond to columns indexed by $\mathcal{I} \setminus \mathcal{S}_{\text{opt}}$. Then if $\boldsymbol{\eta} = 0$ and

$$M_{\text{OMP}} = \|\tilde{\mathbf{H}}^{\dagger} \bar{\mathbf{H}}\|_{1,1} < 1, \quad (10)$$

OMP recovers support of β exactly in k iterations. This condition is called Exact Recovery Condition (ERC). Similar results can be established for OLS. In particular, the following proposition holds for $\boldsymbol{\eta} = 0$.

Proposition 1. Let $\Phi_{S_i} = [\mathbf{b}_j^{S_i}] \in \mathbb{R}^{n \times (k-i)}$, $j \in \mathcal{S}_{opt} \setminus S_i$ and $\Psi_{S_i} = [\mathbf{b}_j^{S_i}] \in \mathbb{R}^{n \times (m-k)}$, $j \in \mathcal{T} \setminus \mathcal{S}_{opt}$. Suppose OLS identified true columns in first i iterations. If

$$M_{i+1} = \|\Phi_{S_i}^\dagger \Psi_{S_i}\|_{1,1} < 1, \quad (11)$$

OLS chooses a true column in $(i+1)^{st}$ iteration.

The condition $M_i < 1$ is the ERC of OLS at i^{th} iteration. Here we extend this results to the case that the measurements are perturbed with additive noise where $\|\boldsymbol{\eta}\|_2 \leq \epsilon_\eta$. The following theorem summarizes our main results.

Theorem 3.1. Suppose $\|\boldsymbol{\eta}\|_2 \leq \epsilon_\eta$ and that OLS has chosen true columns in first i iteration. Then OLS selects an index from \mathcal{S}_{opt} at $(i+1)^{th}$ iteration if $M_{i+1} < 1$ and

$$\min_j |\bar{\beta}_j| > \sigma_{\min}(\bar{\mathbf{H}})\epsilon_\eta + \frac{\epsilon_\eta}{(1 - M_{i+1})\sigma_{\min}^2(\bar{\mathbf{H}})}. \quad (12)$$

Proof. Proof follows an inductive argument. First, without loss of generality, assume columns of \mathbf{H} are normalized and all nonzero components of β are in first k locations. This implies that \mathbf{H} can be written in form of $\mathbf{H} = [\bar{\mathbf{H}}, \tilde{\mathbf{H}}]$. Assume OLS has selected columns from \mathcal{S}_{opt} in first i iterations. It then follows from 4 that

$$\rho(\mathbf{r}_i) = \frac{\|\Psi_{S_i}^\top \mathbf{r}_i\|_\infty}{\|\Phi_{S_i}^\top \mathbf{r}_i\|_\infty} < 1. \quad (13)$$

is a sufficient condition so that OLS selects a true column at next iteration. However,

$$\begin{aligned} \rho(\mathbf{r}_i) &\stackrel{(a)}{\leq} \frac{\|\Psi_{S_i}^\top \boldsymbol{\eta}^\perp + \Psi_{S_i}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty}{\|\Phi_{S_i}^\top \boldsymbol{\eta}^\perp + \Phi_{S_i}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty} \\ &\stackrel{(b)}{\leq} \frac{\|\Psi_{S_i}^\top \boldsymbol{\eta}^\perp\|_\infty + \|\Psi_{S_i}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty}{\|\Phi_{S_i}^\top \boldsymbol{\eta}^\perp + \Phi_{S_i}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty} \\ &\stackrel{(c)}{\leq} \frac{\|\Psi_{S_i}^\top \boldsymbol{\eta}^\perp\|_\infty + \|\Psi_{S_i}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty}{\|\Phi_{S_i}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty} \end{aligned} \quad (14)$$

where (a) is by the equivalence definition of \mathbf{r}_i in Lemma 2.1, (b) follows from triangle inequality, and (c) is due to the fact that $\boldsymbol{\eta}^\perp$ is orthogonal to \mathcal{L}_k . Let $\mathbf{u} = \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}$. Hence, we may calculate that

$$\begin{aligned} \frac{\|\Psi_{S_i}^\top \mathbf{u}\|_\infty}{\|\Phi_{S_i}^\top \mathbf{u}\|_\infty} &= \frac{\|\Psi_{S_i}^\top (\Phi_{S_i}^\dagger)^\top \Phi_{S_i}^\top \mathbf{u}\|_\infty}{\|\Phi_{S_i}^\top \mathbf{u}\|_\infty} \\ &\leq \|\Psi_{S_i}^\top (\Phi_{S_i}^\dagger)^\top\|_{\infty, \infty} \\ &= \|\Phi_{S_i}^\dagger \Psi_{S_i}\|_{1,1} \end{aligned} \quad (15)$$

owing to the fact that \mathbf{u} lies in $\mathcal{L}_k \setminus \mathcal{L}_i$, and the relation between $\|\cdot\|_{1,1}$ and $\|\cdot\|_{\infty, \infty}$. Therefore, by definition of M_i in Proposition 1

$$\rho(\mathbf{r}_i) \leq M_{i+1} + \frac{\|\Psi_{S_i}^\top \boldsymbol{\eta}^\perp\|_\infty}{\|\Phi_{S_i}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty} \quad (16)$$

Now, observe that applying Lemma 2.1 along with the fact that \mathbf{P}_i^\perp is a projection matrix delivers

$$\|\bar{\mathbf{H}}_{i^c} \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_2 \geq \sigma_{\min}^2(\bar{\mathbf{H}}) \|\mathbf{z}_{i^c}\|_2. \quad (17)$$

Consequently, noting $\max_{j \in \mathcal{S}_{opt} \setminus S_i} \|\mathbf{P}_i^\perp \mathbf{a}_j\|_2 = 1$, one may continue to obtain

$$\begin{aligned} \rho(\mathbf{r}_i) &\leq M_{i+1} + \frac{\|\Psi_{S_i}^\top \boldsymbol{\eta}^\perp\|_\infty}{\|\bar{\mathbf{H}}^\top \mathbf{P}_i^\perp \bar{\mathbf{H}}_{i^c} \mathbf{z}_{i^c}\|_\infty} \\ &\stackrel{(a)}{\leq} M_{i+1} + \frac{\sqrt{k-i} \|\Psi_{S_i}^\top \boldsymbol{\eta}^\perp\|_\infty}{\sigma_{\min}^2(\bar{\mathbf{H}}) \|\mathbf{z}_{i^c}\|_2} \\ &\stackrel{(b)}{\leq} M_{i+1} + \frac{\sqrt{k-i} \epsilon_\eta}{\sigma_{\min}^2(\bar{\mathbf{H}}) \|\mathbf{z}_{i^c}\|_2} \end{aligned} \quad (18)$$

where (a) is by 17 and the fact that $\mathbf{z}_{i^c} \in \mathbb{R}^{k-i}$, and (b) follows from $\|\boldsymbol{\eta}\|_2 \leq \epsilon_\eta$ and the fact that columns of Ψ_{S_i} have unit l_2 norm. Define $\beta_{\min} = \min_j |\bar{\beta}_j|$ and $\mathbf{z}_{\min} = \min_j \|\mathbf{z}_j\|_2$. It is easy to check $\mathbf{z}_{\min} \geq \beta_{\min} - \|\mathbf{w}\|_2$. Hence, one may obtain

$$\begin{aligned} \|\mathbf{z}_{i^c}\|_2 &\geq \sqrt{k-i} \mathbf{z}_{\min} \\ &\geq \sqrt{k-i} (\beta_{\min} - \|\mathbf{w}\|_2) \\ &= \sqrt{k-i} (\beta_{\min} - \|\bar{\mathbf{H}}^\dagger \boldsymbol{\eta}\|_2) \\ &\geq \sqrt{k-i} (\beta_{\min} - \sigma_{\max}(\bar{\mathbf{H}}^\dagger) \|\boldsymbol{\eta}\|_2) \\ &= \sqrt{k-i} (\beta_{\min} - \sigma_{\min}(\bar{\mathbf{H}}) \epsilon_\eta). \end{aligned} \quad (19)$$

where we imposed $\beta_{\min} > \sigma_{\min}(\bar{\mathbf{H}}) \epsilon_\eta$. Combine 18 and 19 to reach

$$\rho(\mathbf{r}_i) \leq M_{i+1} + \frac{\epsilon_\eta}{\sigma_{\min}^2(\bar{\mathbf{H}}) (\beta_{\min} - \sigma_{\min}(\bar{\mathbf{H}}) \epsilon_\eta)} \quad (20)$$

Therefore, since $M_{i+1} < 1$ by assumption, condition 12² is sufficient for $\rho(\mathbf{r}_i) < 1$ whence OLS selects a true column in iteration $(i+1)^{st}$. This completes the proof. ■

Remark 1: Note that Theorem 3.1 can be tailored to obtain ERC-based conditions for OMP by replacing M_{i+1} in 12 with M_{OMP} . Here we compare 12 with the result of Proposition 1 in [15]. If

$$\sigma_{\min}^3(\bar{\mathbf{H}}) < 1/(1 - M_{OMP}), \quad (21)$$

Proposition 1 in [15] requires a more restrictive condition than 12 on the small element of $\bar{\beta}$. However, for nearly all \mathbf{H} that are of interest in compressed sensing applications, 21 is satisfied with high probability as M_{OMP} is inversely proportional to k . Therefore, Theorem 3.1 ameliorates the existing ERC-based results for OMP as well as providing new sufficient conditions for OLS.

²It should be noted that 12 does not conflict with our restriction of $\beta_{\min} > \sigma_{\min}(\bar{\mathbf{H}}) \epsilon_\eta$.

4. EXPLOITING RANDOMNESS IN MEASUREMENTS

When \mathbf{H} is drawn at random according to $\mathcal{N}(0, 1/n)$ or $\mathcal{B}(\frac{1}{2}, \pm \frac{1}{\sqrt{n}})$, concentration of measure inequalities such as that of Lemma 2.3 hold. In particular, singular values of \mathbf{H} are with high probability concentrated around 1. Therefore, one may exploit these properties to establish probabilistic and perhaps pragmatic guarantees for performance of a sparse reconstruction algorithm. Theorem 4.1 below, states that for these matrices, OLS is capable of exact recovery of sparse signals with high probability if elements of β are sufficiently larger than noise.

Theorem 4.1. *Suppose β is an arbitrary sparse vector with sparsity level k in \mathbb{R}^m . Choose a random matrix $\mathbf{H} \in \mathbb{R}^{n \times m}$ such that its entries are drawn uniformly and independently from $\mathcal{N}(0, 1/n)$ or $\mathcal{B}(\frac{1}{2}, \pm \frac{1}{\sqrt{n}})$. Fix $0 < \epsilon < 1$, $0 < \delta < 1$, and $t > 0$. Given the noisy measurements $\mathbf{y} = \mathbf{H}\beta + \boldsymbol{\eta}$ where $\boldsymbol{\eta}$ is independent of \mathbf{H} and β and $\|\boldsymbol{\eta}\|_2 \leq \epsilon_\eta$, if $\min_{\beta_j \neq 0} |\beta_j| \geq (1 + \delta + t)\epsilon_\eta$, OLS recovers β in k iterations with probability of success exceeding*

$$\geq \left(1 - 2e^{-(n-k+1)c_0(\epsilon)}\right)^2 \left(1 - 2\left(\frac{12}{\delta}\right)^k e^{-nc_0(\frac{\delta}{2})}\right) \left(1 - 2 \sum_{i=0}^{k-1} e^{-\frac{n \frac{1-\epsilon}{1+\epsilon} (1-\delta)^4}{k \left[\frac{1}{(k-i)t^2} + (1+\delta)^2\right]}}\right)^{m-k}. \quad (22)$$

Proof. In proof of Theorem 3.1 we discussed that 13 is a sufficient condition for OLS to select a true column at next iteration. Therefore, if Σ denotes the event that OLS succeeds, then $\Pr\{\Sigma\} \geq \Pr\{\max_i \rho(\mathbf{r}_i) < 1\}$. Recall the idempotent property for \mathbf{P}_i^\perp , i.e.,

$$\mathbf{P}_i^\perp = \mathbf{P}_i^{\perp 2} = \mathbf{P}_i^{\perp \top} \quad (23)$$

Employing Lemma 2.3, along with 23 delivers

$$\rho(\mathbf{r}_i) \leq \frac{1}{c_1(\epsilon)} \frac{\|\tilde{\mathbf{H}}^\top \mathbf{r}_i\|_\infty}{\|\tilde{\mathbf{H}}^\top \mathbf{r}_i\|_\infty} \quad (24)$$

with probability exceeding $p_1 = \left(1 - 2e^{-(n-k+1)c_0(\epsilon)}\right)^2$ for $0 \leq i < k$ where $c_1(\epsilon) = \sqrt{\frac{1-\epsilon}{1+\epsilon}}$. Following the framework in [17], a simple norm inequality and the fact that $\tilde{\mathbf{H}}^\top \mathbf{r}_i$ has at most k nonzero entries results in

$$\rho(\mathbf{r}_i) \leq \frac{\sqrt{k}}{c_1(\epsilon)} \|\tilde{\mathbf{H}}^\top \tilde{\mathbf{r}}_i\|_\infty \quad (25)$$

where $\tilde{\mathbf{r}}_i = \mathbf{r}_i / \|\tilde{\mathbf{H}}^\top \mathbf{r}_i\|_2$. Consequently, We examine an upper bound for $\tilde{\mathbf{r}}_i$. Employ Lemma 2.1 and Lemma 2.2 in definition of $\tilde{\mathbf{r}}_i$ to reach

$$\|\tilde{\mathbf{r}}_i\|_2 \leq \frac{\left[\|\boldsymbol{\eta}^\perp\|_2^2 / \|\mathbf{z}_{ic}\|_2^2 + \sigma_{\max}^2(\tilde{\mathbf{H}})\right]^{\frac{1}{2}}}{\sigma_{\min}^2(\tilde{\mathbf{H}})}. \quad (26)$$

It is shown in [26] that for any $0 < \delta < 1$,

$$\Pr\{1 - \delta \leq \sigma_{\min}(\tilde{\mathbf{H}}) \leq 1 + \delta\} \geq 1 - 2\left(\frac{12}{\delta}\right)^k e^{-nc_0(\frac{\delta}{2})}. \quad (27)$$

Call the term on the right hand side of 27 p_2 . Combining 26, 27, 19, and the fact that $\|\boldsymbol{\eta}^\perp\|_2 \leq \|\boldsymbol{\eta}\|_2 \leq \epsilon_\eta$ furnishes

$$\|\tilde{\mathbf{r}}_i\|_2 \leq \frac{\left[\frac{\epsilon_\eta^2}{(k-i)(\beta_{\min} - (1+\delta)\epsilon_\eta)^2} + (1+\delta)^2\right]^{\frac{1}{2}}}{(1-\delta)^2} \quad (28)$$

with probability exceeding p_2 . Thus, imposing the constraint $\beta_{\min} \geq (1 + \delta + t)\epsilon_\eta$ for any $t > 0$ establishes

$$\|\tilde{\mathbf{r}}_i\|_2 \leq \frac{\left[\frac{1}{(k-i)t^2} + (1+\delta)^2\right]^{\frac{1}{2}}}{(1-\delta)^2}. \quad (29)$$

Using the independence assumption of columns of $\tilde{\mathbf{H}}$, the fact that $\{\tilde{\mathbf{r}}_i\}_{i=0}^{k-1}$ are bounded with probability higher than p_2 and are statistically independent of $\tilde{\mathbf{H}}$, and applying Boole's inequality or Hoeffding inequality³ we reach 22 which completes the proof. ■

Remark 2: If we define $\text{SNR} = \frac{\|\mathbf{H}\beta\|_2^2}{\|\boldsymbol{\eta}\|_2^2}$, the condition $\min_{\beta_j \neq 0} |\beta_j| \geq (1 + \delta + t)\epsilon_\eta$ implies

$$\text{SNR} \approx k(1 + \delta + t)^2, \quad (30)$$

which suggests that for exact support recovery via OLS, SNR should scale linearly with sparsity level.

Remark 3: Note that when $k \rightarrow \infty$ (so do m , and n), p_1 , p_2 , and p_3 overwhelmingly approach 1. Therefore, for these set of parameters, one may assume very small ϵ and δ .

With some numerical estimates on lower bound of success probability which was established in Theorem 4.1, one may establish a lower bound on sufficient number of measurements for exact support recovery. Specifically, if conditions within Theorem 4.1 are satisfied, for any $0 < \gamma < 1$, there exist positive constants C_1 , C_2 , and C_3 which are independent of γ , n , m , and k such that OLS succeeds with $\Pr\{\Sigma\} \geq 1 - \gamma^2$ if $n \geq \max\{\frac{2}{C_1} k \log \frac{m}{\gamma}, C_2 k + \log \frac{12}{\gamma^2} / C_3\}$. Hence, OLS recovers k -sparse $\beta \in \mathbb{R}^m$ if the number of measurements grow linearly in k and logarithmically in m .

5. CONCLUSION

In this paper, we established sufficient conditions for exact support recovery via Orthogonal Least-Squares (OLS) in noisy setting. In particular, if ERC for OLS holds, and SNR is adequately high, OLS recovers all true indices. We also showed that for Gaussian and Bernoulli coefficient matrices, OLS is with high probability guaranteed to reconstruct any sparse signal from a low number of noisy random linear measurements if its elements are too small.

³Depending on whether $\mathbf{H} \sim \mathcal{N}(0, \frac{1}{n})$ or $\mathbf{H} \sim \mathcal{B}(\frac{1}{2}, \pm \frac{1}{\sqrt{n}})$.

6. REFERENCES

- [1] Michael E Tipping, "Sparse bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.
- [2] David L Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] Cecilia Carbonelli, Satish Vedantam, and Urbashi Mitra, "Sparse channel estimation with zero tap detection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1743–1763, 2007.
- [4] Somsubhra Barik and Haris Vikalo, "Sparsity-aware sphere decoding: algorithms and complexity analysis," *Signal Processing, IEEE Transactions on*, vol. 62, no. 9, pp. 2212–2225, 2014.
- [5] Farzad Parvaresh, Haris Vikalo, Sidhant Misra, and Babak Hassibi, "Recovering sparse signals using sparse measurement matrices in compressed dna microarrays," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 2, no. 3, pp. 275–285, 2008.
- [6] Michael Lustig, David Donoho, and John M Pauly, "Sparse mri: The application of compressed sensing for rapid mr imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [7] Michael Elad, Mario AT Figueiredo, and Yi Ma, "On the role of sparse and redundant representations in image processing," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 972–982, 2010.
- [8] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2790–2797.
- [9] Yagyensh Chandra Pati, Ramin Rezaifar, and PS Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE, 1993, pp. 40–44.
- [10] David L Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1094–1121, 2012.
- [11] Wei Dai and Olga Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, 2009.
- [12] Jian Wang, Seokbeop Kwon, and Byonghyo Shim, "Generalized orthogonal matching pursuit," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6202–6216, 2012.
- [13] Qun Mo and Yi Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 58, no. 6, pp. 3654–3656, 2012.
- [14] Joel A Tropp, "Greed is good: Algorithmic results for sparse approximation," *Information Theory, IEEE Transactions on*, vol. 50, no. 10, pp. 2231–2242, 2004.
- [15] T Tony Cai and Lie Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *Information Theory, IEEE Transactions on*, vol. 57, no. 7, pp. 4680–4688, 2011.
- [16] Abhimanyu Das and David Kempe, "Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 1057–1064.
- [17] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [18] Alyson K Fletcher and Sundeep Rangan, "Orthogonal matching pursuit: A brownian motion analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1010–1021, 2012.
- [19] Sheng Chen, Stephen A Billings, and Wan Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of control*, vol. 50, no. 5, pp. 1873–1896, 1989.
- [20] Charles Soussen, Rémi Gribonval, Jérôme Idier, and Cédric Herzet, "Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares," *Information Theory, IEEE Transactions on*, vol. 59, no. 5, pp. 3158–3174, 2013.
- [21] Shay Maymon and Yonina C Eldar, "The viterbi algorithm for subset selection," *IEEE Signal Processing Letters*, vol. 22, no. 5, pp. 524–528, 2015.
- [22] Abolfazl Hashemi and Haris Vikalo, "Sampling requirements and accelerated schemes for sparse linear regression with orthogonal least-squares," *preprint*, 2016.
- [23] Cédric Herzet, Angélique Drémeau, and Charles Soussen, "Relaxed recovery conditions for omp/ols by exploiting both coherence and decay," *Information Theory, IEEE Transactions on*, vol. 62, no. 1, pp. 459–470, 2016.
- [24] Cédric Herzet, Charles Soussen, Jérôme Idier, and Rémi Gribonval, "Exact recovery conditions for sparse representations with partial support information," *Information Theory, IEEE Transactions on*, vol. 59, no. 11, pp. 7509–7524, 2013.
- [25] Abolfazl Hashemi and Haris Vikalo, "Sparse linear regression via generalized orthogonal least-squares," *arXiv preprint arXiv:1602.06916*, 2016.
- [26] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.